DOCUMENT RESUME

ED 441 829		TM 030 864
AUTHOR	Mecklin, Christopher J.; Mundfrom, Daniel J.	
TITLE	Comparing the Power of Classical and Newer Tests of Multivariate Normality.	
PUB DATE	2000-04-27	
NOTE	37p.; Paper presented at the Annual Meeting of the American Educational Research Association (81st, New Orleans, LA, April 24-28, 2000).	
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)	
EDRS PRICE	MF01/PC02 Plus Postage.	
DESCRIPTORS	Monte Carlo Methods; *Multivariate Analysis; Simulation	
IDENTIFIERS	*Normality Tests; Power (Statistics)	

ABSTRACT

Many multivariate statistical methods call upon the assumption of multivariate normality. However, many applied researchers fail to test this assumption. This omission could be due to ignorance of the existence of tests of multivariate normality or confusion about which test to use. Although at least 50 tests of multivariate normality exist, relatively little is known about the power of these procedures. The purpose of this study was to examine the power of 13 promising tests of multivariate normality under a variety of conditions. Monte Carlo simulations were used to generate 10,000 data sets from many multivariate distributions, including the multivariate normal distribution, normal mixtures, elliptically contoured distributions, and heavily skewed distributions. The test statistic for each procedure was calculated and compared with the appropriate critical value. The number of rejections of the null hypothesis of multivariate normality was tabled for each situation. No single test was found to be the most powerful in all situations. The use of the Henze-Zirkler (N. Henze and B. Zirkler, 1990) test is recommended for a formal test of the null hypothesis of multivariate normality. The use of supplementary procedures such as K. Mardia's measures of skewness and kurtosis and the chi-square or beta plot is also recommended for diagnosing the cause of the non-normality. (Contains 11 tables and 48 references.) (Author/SLD)

Running Head: MULTIVARIATE NORMALITY

Comparing the Power of Classical and Newer Tests of Multivariate Normality

Christopher J. Mecklin and Daniel J. Mundfrom

University of Northern Colorado

Paper presented at the annual meeting of the American Educational Research Association

New Orleans, LA, April 27, 2000

Abstract

Many multivariate statistical methods call upon the assumption of multivariate normality. However, many applied researchers fail to test this assumption. This omission could be due to either ignorance of the existence of tests of multivariate normality or confusion about which test to use. Although at least 50 tests of multivariate normality exist, relatively little is known about the power of these procedures. The purpose of this study was to examine the power of 13 promising tests of multivariate normality under a variety of conditions. Monte Carlo simulations were used to generate 10,000 data sets from many multivariate distributions, including the multivariate normal distribution, normal mixtures, elliptically contoured distributions, and heavily skewed distributions. The test statistic for each procedure was calculated and compared with the appropriate critical value. The number of rejections of the null hypothesis of multivariate normality was tabled for each situation. No single test was found to be the most powerful in all situations. The use of the Henze-Zirkler test is recommended for a formal test of the null hypothesis of multivariate normality. The use of supplementary procedures such as Mardia's measures of skewness and kurtosis and the chi-square or beta plot is also recommended for diagnosing the cause of the non-normality.

Introduction

Multivariate methodology is vital to not only avoid Type I error rate inflation, but also to honor the reality that most effects have multiple causes and multiple consequences (Thompson, 1996). Also, it is well known that many multivariate statistical methods, including MANOVA, discriminant analysis, and canonical correlation, call upon the assumption of multivariate normality. According to Anderson (1984, pg. 3), "A major reason for basing statistical analysis on the normal distribution is that this probabilistic model approximates well the distribution of continuous measurements in many sampled populations."

The performance of many multivariate methods is affected by certain deviations from normality (Looney, 1995). Hypothesis tests involving mean vectors are more sensitive to skewness, while tests involving variance-covariance matrices are more sensitive to kurtosis (Mardia, Kent, & Bibby, 1979; DeCarlo, 1997). However, the assumption of multivariate normality often goes untested (Baxter, 1997). Horswell (1990, pg. 162) declared that these tests are "largely academic curiosities, seldom used by practicing statisticians." Possible explanations for this omission by practitioners include (Looney, 1995):

1. The practitioner is unaware of the existence of tests of multivariate normality.
2. Convenient software for calculating the test statistic or p-value for a test of multivariate normality is not readily available.
3. Even if software is used to calculate the test statistic, a special table may be necessary to approximate the p-value of the test.

4. The practitioner may not want to use a procedure when little is known about the statistical power of the test.
5. The practitioner is reluctant to test for multivariate normality because he is unsure of how to proceed if non-normality is detected.

An extensive literature exists regarding the testing of multivariate normality. One reason that this literature exists is the fact that no single procedure can be uniformly most powerful against all possible alternatives, or departures, from normality. At least 50 different procedures have been proposed for this problem. Since the possible variations from normality are endless, Andrews, Gnanadesikan, and Warner (1973, pg. 95) warned "seeking a single best method would seem to be neither pragmatically sensible nor necessary."

The main purpose of this study is to pinpoint which procedures for testing multivariate normality are effective against a wide range of non-normal alternatives. The procedure(s) that are identified as effective could be used by a researcher even and especially when the true distribution of the population is not known *a priori*. More specifically, 13 different tests of multivariate normality were compared in a Monte Carlo simulation study against many different distributions, ranging from the multivariate normal to severe departures from normality. The research questions listed below all pertain to the power of the following "classical" (such as Mardia's measures) and "newer" tests of multivariate normality:

- Mardia's test of multivariate skewness (1970)
- Mardia's test of multivariate kurtosis (1970)
- Hawkins' test (1981)

- Koziol's test (1982)
- The Mardia-Foster omnibus test (1983)
- Royston's test (1983)
- The Paulson-Roohan-Sullo test (1987)
- The Henze-Zirkler test (1990)
- The Mardia-Kent omnibus test (1991)
- The Romeu-Ozturk test (1993)
- Singh's classical and robust tests (1993)
- The Mudholkar-Srivastava-Lin test (1995)

The research questions to be addressed on this body of procedures for testing multivariate normality are:

1. Do these tests reject the null hypothesis of multivariate normality at the stated alpha level with data from a multivariate normal population?
2. Which of these tests are most powerful against multivariate normal mixtures?
3. Which of these tests are most powerful against elliptically contoured distributions?
4. Which of these tests are most powerful against heavily skewed distributions?
5. Which of these tests are most powerful when faced with a non-normal multivariate distribution whose marginals are normal?
6. Which of these tests are most powerful when faced with a non-normal multivariate distribution that has multivariate normal values for skewness and kurtosis?

Theoretical Framework

In general, analyses based upon variance-covariance matrices can be seriously affected by the kurtosis of the distribution while analyses that involve the mean vectors are more sensitive to skewness (Mardia, Kent, & Bibby, 1979; DeCarlo, 1997). But in many situations, we are either performing several different multivariate procedures upon a data set or a procedure that involves hypotheses and/or assumptions about both location and dispersion. Thus, generally we are concerned about deviations in terms of skewness, kurtosis, or both simultaneously, and thus require a flexible method of testing multivariate normality.

There is no shortage of proposed methods for assessing multivariate normality. A current review of the literature has revealed that at least 50 procedures for testing multivariate normality exist. Despite the abundance of methods, Rencher (1995) commented that since multivariate normality is not as straightforward as univariate normality, the "state of the art" is not as refined. Several reviews of the different methods exist (see Andrews, Gnanadesikan, & Warner, 1973; Gnanadesikan, 1977; Mardia, 1980; Koziol, 1986; and Looney, 1995), but none are completely comprehensive. When compared to the amount of research available in developing tests of multivariate normality, relatively little work has been done in evaluating the quality and power of these procedures. Examples of studies comparing the power of tests of multivariate normality are Ward (1988), Horswell (1990), Horswell & Looney (1992), Romeu & Ozturk (1993), Young, Seaman, & Seaman (1995), and Bogdan (1999). None of these studies is completely comprehensive and most were deliberately restricted in scope to a limited category of tests.

Much of multivariate statistics consists of extensions of univariate methods to the general case. Testing the goodness of fit of a data set to the multivariate normal distribution is no exception. Most of the available multivariate normality testing procedures are extensions of simpler tests of univariate normality. Thus, a large percentage of multivariate normality tests are either based on graphical plots, measures of skewness and/or kurtosis, or goodness-of-fit procedures. Unfortunately, few of these tests are truly formal, in the sense that both the null distribution of the test statistic has been found and the consistency of the test has been established (Koziol, 1983; Bogdan, 1999). In the words of Baringhaus and Henze (1988, pg. 399) and of Csorgo (1989, pg. 108), there are few "genuine" tests of multivariate normality.

In the review of the literature, four categories of tests of multivariate normality were found which could classify virtually all of the available procedures. These categories are:

1. Graphical and Correlational Approaches
2. Skewness and Kurtosis Approaches
3. Goodness-of-fit Approaches
4. Consistent Approaches

A very common informal approach to univariate normality is to construct a normal probability plot or a quantile-quantile (Q-Q) plot. Normality is indicated if this plot is linear. A more formal hypothesis test can be based upon the correlation of the Q-Q plot. This time-honored approach to assessing normality was extended to the multivariate situation by Healy (1968) with the chi-square plot. In this classical procedure, the squared Mahalanobis distances are ordered and plotted against approximate expected order statistics from the chi-square distribution. Gnanadesikan and Kettenring (1972) were the first to note that the exact marginal distribution of the squared Mahalanobis distances is a multiple of a beta distribution. For a purely visual inspection, the difference between the chi-square or beta plot is insignificant. However, this difference is more crucial in developing a formal hypothesis test for multivariate normality.

Singh (1993) developed two different tests based upon the correlation of the beta plot. One version of the test used the standard classical estimators for the mean vector and the variance-covariance matrix for calculating the Mahalanobis distances. The second version of this test used robust M-estimators (Maronna, 1976) for the mean vector and variance-covariance matrix. The robust version of this test was developed to alleviate the fact that multivariate outliers would greatly influence the calculated value of the Mahalanobis distances and thus the beta plot's correlation coefficient. Singh's procedures have not been previously assessed in a comprehensive power study.

Mardia's (1970) introduction of multivariate measures of skewness and kurtosis was another seminal paper in the field of multivariate normality testing. Mardia derived affine invariant extensions for skewness and kurtosis. The parameters for multivariate skewness and kurtosis are denoted $\beta_{1,p}$ and $\beta_{2,p}$, respectively. For the multivariate normal distribution, $\beta_{1,p} = 0$ and $\beta_{2,p} = p(p+2)$. Mardia determined that a function of the multivariate skewness is asymptotically distributed as a chi-square random variable with $\frac{p(p+1)(p+2)}{6}$ degrees of freedom and a function of the multivariate kurtosis is asymptotically distributed as a standard normal random variable. Mardia exploited this to develop two tests for multivariate normality.

Mardia's procedures, particularly his test based on multivariate kurtosis, are probably the most used tests of multivariate normality. Mardia's measures are available in SAS in PROC CALIS or PROC MODEL and the kurtosis measure is available in several structural equation modeling packages. Previous research (Ward, 1988; Horswell, 1990; Horswell & Looney, 1992; Romeu & Ozturk, 1993; Bogdan, 1999) has indicated that Mardia's procedures, particularly the skewness test, are among the best available tests. It is inconceivable for any comprehensive study of the power of tests of multivariate normality to not consider Mardia's measures. Thus, both were considered in this study.

Many efforts have been made to construct a single "omnibus" test statistic that combines elements of both skewness and kurtosis. Mardia and Foster (1983) derived six possible test statistics. A statistic denoted S_w^2, which used the Wilson-Hilferty approximation (Bain & Englehardt, 1992) to derive an omnibus statistic with an asymptotic chi-square distribution with two degrees of freedom, was found to be not powerful by Ward (1988) or Horswell and Looney (1992). However, an alternative statistic, C_w^2, which factored in the covariance between multivariate skewness and kurtosis, has not been previously studied. A more recent and promising statistic is an omnibus statistic derived using Rao scores (Mardia & Kent, 1991). Both the C_w^2 test and the Mardia-Kent test were considered.

A multitude of researchers have extended univariate goodness-of-fit procedures to the general multivariate case in order to develop a test for multivariate normality. Many goodness-of-fit procedures, such as the Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling tests, are based on the empirical distribution function. A simpler but less powerful technique is the familiar chi-square test. Although many promising tests of multivariate normality fall into this category, Rencher (1995) has criticized this approach as unrealistic due to the inherent "sparseness" in multivariate data.

Hawkins (1981) proposed an extension of the Anderson-Darling statistic to test a multivariate data set for both normality and homoscedasticity. Paulson, Roohan, and Sullo (1987) proposed a similar test. Both procedures convert the squared Mahalanobis distances into a statistic that will have a uniform distribution if the multivariate normal distribution holds. Their inclusion in this study is due to encouraging results reported by Ward (1988), Romeu and Ozturk (1993), and Young, Seaman, and Seaman (1995).

Koziol (1982) derived his test by extending the Cramer-von Mises test to employ the ordered squared Mahalanobis distances and approximate expected order statistics from the chi-square distribution originally used by Healy (1968). Koziol's test had relatively high power in the comparison studies of Romeu and Ozturk (1993) and Young, Seaman, and Seaman (1995), leading to its consideration in this study.

The Shapiro-Wilk (1965) test is generally considered to be among the best procedures for assessing univariate normality. Thus, it is only natural to extend it to the multivariate case, as done by Royston (1983) and Mudholkar, Srivastava, and Lin (1995). Although both procedures were considered here, Royston's procedure was found to be sensitive to the correlational structure between the variables (Romeu & Ozturk, 1993).

A more recent and less well known goodness-of-fit procedure is the Q_n procedure developed by Ozturk and Dudewicz (1992). Romeu and Ozturk (1993) expanded this procedure to test for multivariate normality. Their own power study indicated that the Cholesky root version of this procedure was one of the most powerful procedures available. The Romeu-Ozturk test was considered in this study.

All of the above procedures have been criticized for their lack of consistency (Baringhaus & Henze, 1988; Csogro, 1989; Henze & Zirkler, 1990; Bogdan, 1999).

Recall that a statistic is *consistent* if it converges in probability to the parameter that it is estimating and is generally considered to be a mathematically desirable property (Hogg & Craig, 1995).

Epps and Pulley (1983) developed a test for univariate normality based upon the empirical characteristic function. Baringhaus and Henze (1988) extended this test to the multivariate case. It was proven to be consistent against all alteratives by Csorgo (1989) and further extended into its current form by Henze and Zirkler (1990).

The Henze-Zirkler test statistic is based on using characteristic functions to measure the distance between the hypothesized function (i.e. multivariate normal) and the observed, or empirical, function. For consistency to hold, this distance must equal zero if and only if the observed data are multivariate normal. Extensive derivations yielded a closed form for the Henze-Zirkler statistic, which has an approximate lognormal distribution.

This test was included in the power study largely due to the consistency of the statistic. The consistency result implies the strong possibility that this test will be very competitive against a wide range of alternatives and that this test is unlikely to have a major weakness. Further, a similar test due to Bowman and Foster (1993) was found to be a good performer by Bogdan (1999). However, Henze (1997) and Henze and Wagner (1997) pointed out that Bowman and Foster's integrated squared error statistic is actually a special case of the Henze-Zirkler test.

Methodology

The purpose of this study was to compare the power of commonly used and promising tests of multivariate normality. Because the null distribution of most test statistics for multivariate normality is intractable, the Monte Carlo technique was used. SAS/IML (Interactive Matrix Language) was used for the simulations. In the simulations, 10000 data sets with sample size n = 25, 50, or 100 and dimension p = 2, 3, 4, or 5 from both the multivariate normal distribution and several non-normal distributions were generated. These limitations were imposed both to keep computing time reasonable and to consider sample sizes that are borderline for multivariate analysis. These small sample sizes are likely to be the situation where the assumption of multivariate normality is most critical to the researcher. The test statistic was calculated for all 10000 data sets and compared to the appropriate critical value in order to estimate the proportion of rejections for each test in each situation.

As a summary, the 13 tests of multivariate normality from four distinct categories to be considered are:

A. Graphical and correlational test

1. Singh's test (1993) of the correlation of the beta plot, utilizing classical estimates of location and dispersion
2. Singh's test (1993) of the correlation of the beta plot, utilizing robust M-estimates of location and dispersion

B. Tests of multivariate skewness and kurtosis

1. Mardia's (1970) test of multivariate skewness
2. Mardia's (1970) test of multivariate kurtosis
3. The Mardia-Foster (1983) omnibus statistic
4. The Mardia-Kent (1991) omnibus statistic

C. Multivariate extensions of univariate goodness-of-fit procedures

1. Koziol's (1982) extension of the Cramer-von Mises test
2. Hawkins' (1981) extension of the Anderson-Darling test
3. The Paulson-Roohan-Sullo (1987) extension of the Anderson-Darling test
4. Royston's (1983) multivariate Shapiro-Wilk test
5. The Mudholkar-Srivastava-Lin (1995) extension of the Shapiro-Wilk test
6. The Romeu-Ozturk (1993) test

D. Consistent tests of multivariate normality that use the empirical characteristic function

1. The Henze-Zirkler (1990) test

In a Monte Carlo study, it is important to choose the alternatives to multivariate normality very carefully. Many past Monte Carlo studies have been criticized for being "wasteful and superfluous" (Hampel, Ronchetti, Rousseeuw, & Stahel, 1986, pg. 6) or "haphazardly selected" (Horswell, 1990, pg. 167). Another limitation of some past simulation studies was to only consider multivariate distributions that were merely composed of marginal components independently and identically distributed from some familiar univariate distribution. However, uncorrelated variables are neither common nor interesting in multivariate analysis (Ward, 1988). Thus, any reasonable study comparing the power of tests of multivariate normality will consider multivariate distributions with correlated components. The multivariate distributions used for this study fulfilled this criterion and were generated with algorithms found in Johnson (1987).

The first distribution to be considered is the multivariate normal distribution itself. The tests are to be compared against the normal distribution for two reasons: to serve as a check that the algorithms for calculating the test statistics were programmed correctly and to make sure that the tests only reject normality (and thus make a Type I error) at approximately the nominal alpha level.

To simulate the situation of sampled subjects coming from two distinct normal populations, various normal mixtures were considered. Three levels of mixing (or contamination) were considered: $\widetilde{p} = 0.9$, $\widetilde{p} = 0.788675$, and $\widetilde{p} = 0.5$. The first choice indicated mild contamination and is skewed and leptokurtic; the second choice indicates moderate contamination and is skewed and mesokurtic; and the third choice indicates severe contamination and is symmetric and platykurtic (Mardia, Kent, & Bibby, 1979; Horswell, 1990). Further, the second choice, which is a non-normal distribution with normal kurtosis, has been shown by Henze (1994) to be an alternative where Mardia's tests are not consistent and have low power.

Elliptically contoured distributions are symmetric distributions that have contours of equal density that have an elliptical shape (Johnson, 1987; Rencher, 1998). These distributions are closely related to the normal distribution and are mild deviations from normality. In fact, the multivariate normal distribution is a special case of an elliptical distribution. The elliptical distributions considered in this study were the multivariate uniform, which is highly non-normal due to platykurtosis (Romeu & Ozturk, 1993), two members of the Pearson Type II family, and the multivariate t (with 10 degrees of freedom) and Cauchy distributions, which are members of the Pearson Type VII family and are very close to normality.

More severe departures from normality are seen in distributions that fall outside of the elliptically contoured family and thus have skewness. It is expected that tests of multivariate normality would have very high power against this class of distributions. The two examples of heavily skewed distributions considered in this study were the multivariate chi-square and the multivariate lognormal. Both of these distributions also exhibit non-normal kurtosis.

A theoretically interesting departure from multivariate normality is the situation where the univariate marginal distributions are normal but the joint distribution is not. This situation is impossible to detect using only univariate methods and is challenging for even the multivariate procedures. An example of a multivariate distribution that fits this description is a member of the Knintchine family of distributions (Johnson, 1987; Horswell, 1990) and was used in this study to answer the fifth research question.

Another theoretically interesting case that is difficult to detect is a non-normal distribution that has the same values for multivariate skewness and kurtosis as the normal distribution. The "Generalized Exponential Power" family of distributions has this property (Johnson, 1987; Horswell, 1990). A member of this family was used to answer the last research question.

Results

The first distribution simulated was the multivariate normal distribution. In this case, the null hypothesis is true, so each test should reject at about the 5% level. A rejection rate far above the 5% level would indicate a problem with the Type I error rate. The other distributions represent various deviations from multivariate normality, ranging from mild to severe. In these cases, the null hypothesis is false and should be rejected. A low rejection rate, especially in comparison with other tests, would signify a problem with the Type II error rate and the power of the test.

The performance of the 13 tests against the multivariate normal distribution is found in Table 1. One test, the Mardia-Foster, had a very unreliable performance, with the observed power ranging from 0% to 100%. Hawkins' test had an empirical power of 100% when p = 5 and n = 25, while the robust version of Singh's test had very high observed power, ranging from 9.9% to 79.2%. Three other tests (Mudholkar-Srivastava-Lin, Romeu-Ozturk, and Mardia-Kent) had a maximum observed rejection rate of over 10%. An empirical Type I error rate that can be twice the nominal level or greater renders all of these tests as very questionable choices and thus they are not considered further.

INSERT TABLE 1 ABOUT HERE

Fifteen different normal mixture distributions were considered. In general, all of the tests considered had low power against these distributions. As the amount of contamination increased from 10% to 50%, the power increased slightly. The power also increased as the sample size and dimension increased. The most conservative tests in this situation were Mardia's kurtosis, Henze-Zirkler, Koziol, Paulson-Roohan-Sullo, and Singh's classical.

The empirical power of the tests was generally higher for situations where the two different normal distributions had both unequal means and unequal covariances. Tables 2, 3, and 4 give results of the simulations for the normal mixture distributions denoted by $\widetilde{p}N_p(\mu_1,\Sigma_1)+(1-\widetilde{p})N_p(\mu_2,\Sigma_2)$, where $\widetilde{p}$ is a mixing parameter that was set equal to either 0.9, 0.788675, or 0.5, μ_1 is a vector of zeros, μ_2 is a vector of ones, Σ_1 is a correlation matrix with all off-diagonal elements equal to 0.2, and Σ_2 is a correlation matrix with all off-diagonal elements equal to 0.5. Results of the other normal mixtures are similar and are given in Mecklin (2000).

INSERT TABLES 2, 3, AND 4 ABOUT HERE

Five different symmetric distributions from the elliptically contoured family were considered. As one would expect, Mardia's test of multivariate skewness had virtually no power against the multivariate uniform. Singh's classical procedure had power as low as 13.5% against the multivariate uniform. The Henze-Zirkler test and Royston's test were generally powerful, but had minimum power of around 40% when p = 5 and n = 25. Mardia's test of kurtosis had power ranging from 89% to 100% and the tests of Koziol and Paulson-Roohan-Sullo had power of at least 99.8%. Results for the multivariate uniform distribution are given in Table 5. Both versions of the Pearson Type II distribution had extremely similar results, as shown in Mecklin (2000).

INSERT TABLE 5 ABOUT HERE

The Pearson Type VII family of distributions, including the multivariate t and Cauchy, represent mild departures from normality. In general, the rate of rejection was very low for these distributions, generally below 10%. In particular, Mardia's test of kurtosis ranges from only 0.6% to 3.9%. These results are given in Tables 6 and 7.

INSERT TABLES 6 AND 7 ABOUT HERE

Both the multivariate lognormal and chi-square are drastic departures from normality and the power of the tests were very high. For the multivariate chi-square, Mardia's skewness, Royston's test, and Henze-Zirkler all had power of at least 99%, while Mardia's kurtosis had power that dipped as low as 70% when n = 25. As n increased to 100, the power of all procedures was at least 99.9%. The results from the multivariate chi-square are given in Table 8.

INSERT TABLE 8 ABOUT HERE

For the lognormal distribution, empirical power was again very high, with the exception of the erratically performing test of Mardia and Foster. Mardia's kurtosis had the lowest minimum power (82.5%), and the Koziol test had a minimum power of 82.6%. Mardia's skewness, Royston, and Henze-Zirkler all had a minimum power of at least 98%. These results are given in Table 9.

INSERT TABLE 9 ABOUT HERE

Multivariate techniques are crucial for detecting the non-normality of the Knintchine distribution. Since it is symmetric, Mardia's skewness performed very poorly here. Mardia's kurtosis performed modestly when n = 25 (1.8% to 24.8%), considerably better when n = 50 (60.5% to 87.6%), and virtually always rejected normality when n = 100 (99.7% to 100%). The other tests performed similarly to Mardia's kurtosis. These results are in Table 10.

INSERT TABLE 10 ABOUT HERE

The generalized exponential distribution is another theoretically fascinating multivariate distribution, since it has the same values for skewness and kurtosis as the multivariate normal distribution. Not surprisingly, Mardia's tests had virtually no power in this situation. The other procedures also had poor power. These results are in Table 11.

INSERT TABLE 11 ABOUT HERE

Discussion

Many researchers (Bozdogan & Ramirez, 1986; Tsai & Koziol, 1988; Horswell, 1990; Horswell & Looney, 1992; Kariya & George, 1995; Looney, 1995; Mudholkar, Srivastava, & Lin, 1995; Baxter, 1997) have lamented the widespread neglect given to testing the assumption of normality in multivariate analysis. Any attention given towards assessing the assumptions of a statistical procedure is time well spent. Testing for multivariate normality is no exception.

Based upon both previous research (Gnanadesikan, 1977; Koziol, 1986; Ward, 1988; Horswell & Looney, 1992; Romeu & Ozturk, 1993; Looney, 1995; Young, Seaman, & Seaman, 1995; Bogdan, 1999) and the results of this simulation study, no single procedure is the most powerful in all situations. Of the 13 procedures considered in this study, 6 of them (Mardia-Foster, Mardia-Kent, Singh's robust, Mudholkar-Srivastava-Lin, Romeu-Ozturk, Hawkins) had an empirical Type I error rate against the multivariate normal distribution that exceeded 10% in certain circumstances and thus are not recommended.

Of the seven remaining procedures, three of them (Koziol, Mardia's skewness, and Singh's classical) were found to be liberal (i.e. rejected the normal distribution at a rate slightly higher than 5%). Three other procedures (Mardia's kurtosis, Henze-Zirkler, and Paulson-Roohan-Sullo) are conservative, while Royston's test consistently rejected at nearly the nominal level of 5%. The tests of Koziol, Singh, and Paulson-Roohan-Sullo suffer from the additional disadvantage of requiring the use of empirical critical values. The proponents of the use of empirical critical values (Romeu & Ozturk, 1993; Young, Seaman, & Seaman, 1995), even for tests with existing asymptotic null distributions, point out that the asymptotic null distributions are conservative and the use of empirically obtained critical values increases power. However, an advantage of the asymptotic null distributions is that specialized extensive tables are not necessary (Srivastava & Hui, 1987).

Of the seven remaining tests of multivariate normality, five have situations where their power is appreciably less than their competitors. Against heavily skewed deviations from normality, such as the chi-square or lognormal, Mardia's kurtosis, Koziol's test, Singh's test, and the Paulson-Roohan-Sullo test have power as low as 70-80% while Royston's test, the Henze-Zirkler test, and Mardia's skewness have power of virtually 100%. However, Mardia's skewness, as one would expect, has virtually no power against symmetric but non-normal distributions such as the multivariate uniform, t, Cauchy, or Knintchine distributions.

Some procedure for assessing the assumption of multivariate normality should be used. If one is going to rely on only one procedure for this purpose, the Henze-Zirkler test is recommended. This recommendation is based upon both the acceptable Type I error control and power that is either comparable or superior to the other procedures against the entire breadth of considered distributions. Another procedure that had similar power to the Henze-Zirkler and did not suffer from any serious deficiency was Royston's extension of the Shapiro-Wilk test. These empirical results indicate that Royston's test performs at least as well as Henze-Zirkler. However, some theoretical concerns exist for Royston's procedure. Unlike the Henze-Zirkler test, it is not consistent against all alternatives. Further, Royston's test involves a rather ingenious correction for the correlation between the variables in the sample. Unfortunately, this correction has been criticized by Srivastava and Hui (1987) for not being adequately justified and by Romeu and Ozturk (1993) for yielding a test that performs poorly when the variates are highly correlated ($r \approx 0.9$). Then again, data this highly correlated is rarely encountered in educational applications.

Another possibility, originally suggested by Csorgo (1989), would be to use a procedure that is both consistent and powerful to formally test the null hypothesis that the data are from a multivariate normal distribution, and to follow up with less formal procedures if normality was rejected. Based upon the results from this study, the use of the Henze-Zirkler procedure is recommended for conducting the hypothesis test. Since the Henze-Zirkler test statistic does not help in indicating the reason for the rejection of normality, a test rejection should be complemented with graphical procedures such as a chi-square plot and multivariate descriptive statistics such as Mardia's skewness and kurtosis. Based upon these supplemental results, the researcher then could choose the most appropriate next step in the multivariate data analysis.

References

Anderson, T.W. (1984). An introduction to multivariate statistical analysis (2nd Ed.). New York: Wiley.

Andrews, D.F., Gnanadesikan, R., & Warner, J.L. (1973). Methods for assessing multivariate normality. In P.R. Krishnaiah (Ed.), Proceedings of the International Symposium on Multivariate Analysis, Volume 3 (pp. 95-116). New York: Academic Press.

Bain, L.J., & Englehart, M. (1992). Introduction to probability and mathematical statistics. (2nd Ed.). Belmont, CA: Duxbury.

Baringhaus, L., & Henze, N. (1988). A consistent test for multivariate normality based on the empirical characteristic function. Metrika, 35, 339-348.

Baxter, M.J. (1997, April). Testing multivariate normality, with applications to lead isotope data analysis in archaeology. Paper presented at the 25th Anniversary Conference of Computer Applications and Quantitative Methods in Archaeology, Birmingham, UK. Available On-line: http://www.bufau.bham.ac.uk/caa97.baxter/baxter.htm

Bogdan, M. (1999). Data driven smooth tests for bivariate normality. Journal of Multivariate Analysis, 68, 26-52.

Bowman, A.W., & Foster, P.J. (1993). Adaptive smoothing and density based tests of multivariate normality. Journal of the American Statistical Association, 88, 529-537.

Bozdogan, H., & Ramirez, D.E. (1986). Testing for model fit: Asessing Box-Cox transformations of data to near normality. Computational Statistics Quarterly, 3, 127-150.

Csorgo, S. (1989). Consistency of some tests for multivariate normality. Metrika, 36, 107-116.

Epps, T.W., & Pulley, L.B. (1983). A test for normality based on the empirical characteristic function. Biometrika, 70, 723-726.

Gnanadesikan, R. (1977). Methods for statistical data analysis of multivariate observations. New York: Wiley.

Gnanadesikan, R., & Kettenring, J.R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. Biometrics, 28, 81-124.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., & Stahel, W.A. (1985). Robust statistics: The approach based on influence functions. New York: Wiley.

Hawkins, D.M. (1981). A new test for multivariate normality and homoscedasticity. Technometrics, 23, 105-110.

Healy, M.J.R. (1968). Multivariate normal plotting. Applied Statistics, 17, 157-161.

Henze, N. (1994). On Mardia's kurtosis test for multivariate normality. Communications in Statistics: Theory and Methods, 23, 1031-1045.

Henze, N., & Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. Communications in Statistics: Theory and Methods, 19, 3595-3618.

Hogg, R.V., & Craig, A.T. (1995). Introduction to mathematical statistics (5th edition). Englewood Cliffs, NJ: Prentice Hall.

Horswell, R.L. (1990). A monte carlo comparison of tests for multivariate normality based on multivariate skewness and kurtosis. Unpublished doctoral dissertation, Louisiana State University.

Horswell, R.L., & Looney, S.W. (1992). A comparison of tests for multivariate normality that are based on measures of multivariate skewness and kurtosis. Journal of Statistical Computation and Simulation, 42, 21-38.

Huber, P.J. (1981). Robust statistics. New York: Wiley.

Johnson, M.E. (1987). Multivariate statistical simulation. New York: Wiley.

Kariya, T., & George, E.I. (1995). Locally best invariant tests for multivariate normality. Sankhya, 57, 440-451.

Koziol, J.A. (1982). A class of invariant procedures for assessing multivariate normality. Biometrika, 69, 423-427.

Koziol, J.A. (1983). On assessing multivariate normality. Journal of the Royal Statistical Society, Series B, 45, 358-361.

Koziol, J.A. (1986). Assessing multivariate normality: A compendium. Communications in Statistics: Theory and Methods, 15, 2763-2783.

Looney, S.W. (1995). How to use tests for univariate normality to assess multivariate normality. The American Statistician, 49, 64-70.

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. Biometrika, 57, 519-530.

Mardia, K.V. (1980). Tests of univariate and multivariate normality. In P.R. Krishnaiah (Ed.), Handbook of statistics, Volume 1 (pp. 279-320). New York: North Holland.

Mardia, K.V., & Foster, K. (1983). Omnibus tests of multinormality based on skewness and kurtosis. Communications in Statistics, 12, 207-221.

Mardia, K.V., & Kent, J.T. (1991). Rao score tests for goodness of fit and independence. Biometrika, 78, 355-363.

Mardia, K.V., Kent, J.T., & Bibby, J.M. (1979). Multivariate analysis. New York: Academic Press.

Maronna, R.A. (1976). Robust m-estimators of multivariate location and scatter. Annals of Statistics, 4, 51-67.

Mecklin, C.J. (2000). A comparison of the power of classical and newer tests of multivariate normality. Unpublished doctoral dissertation, University of Northern Colorado.

Mudholkar, G.S., Srivastava, D.K., & Lin, C.T. (1995). Some p-variate adaptions of the Shapiro-Wilk test of normality. Communications in Statistics: Theory and Methods, 24, 953-985.

Ozturk, A., & Dudewicz, E.J. (1992). A new statistical goodness-of-fit test based on graphical representation. Biometrical Journal, 34, 403-427.

Paulson, A.S., Roohan, P., & Sullo, P. (1987). Some empirical distribution function tests for multivariate normality. Journal of Statistical Computation and Simulation, 28, 15-30.

Rencher, A.C. (1995). Methods of multivariate analysis. New York: Wiley.

Rencher, A.C. (1998). Multivariate statistical inference and applications. New York: Wiley.

Romeu, J.L., & Ozturk, A. (1993). A comparative study of goodness-of-fit tests for multivariate normality. Journal of Multivariate Analysis, 46, 309-334.

Royston, J.P. (1983). Some techniques for assessing multivariate normality based on the Shapiro-Wilk W. Applied Statistics, 32, 121-133.

Shapiro, S.S., & Wilk, M.B. (1965). An analysis of variance test for normality. Biometrika, 52, 591-611.

Singh, A. (1993). Omnibus robust procedures for assessment of multivariate normality and detection of multivariate outliers. In G.P. Patil and C.R. Rao (Eds.), Multivariate Environmental Statistics (pp.445-488). Amsterdam: North-Holland.

Srivastava, M.S., & Hui, T.K. (1987). On assessing multivariate normality based on Shapiro-Wilk W statistic. Statistics and Probability Letters, 5, 15-18.

Thompson, B. (1996, April). Problems with multivariate normality: Can the multivariate bootstrap help? Paper presented at the annual meeting of the Society for Applied Multivariate Research, Houston (ERIC Reproduction Service No. 398 284).

Tsai, K., & Koziol, J.A. (1988). A correlation type procedure for assessing multivariate normality. Communications in Statistics: Computation and Simulation, 17, 637-651.

Ward, P.J. (1988). Goodness-of-fit tests for multivariate normality. Unpublished doctoral dissertation, University of Alabama.

Young, D.M., Seaman, S.L., & Seaman, J.W. (1995). A comparison of six test statistics for detecting multivariate nonnormality which utilize the multivariate squared-radii statistic. The Texas Journal of Science, 47, 21-38.

Table 1

Empirical Type I Error Rate Against the Multivariate Normal Distribution

MVN	n=25				n=50				n=100			
Test	p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5
Skew	6.3	6.7	6.6	6.2	6.1	5.9	6.1	6.7	5.4	6.1	6.4	6.0
Kurt	0.9	0.5	0.4	1.0	1.8	1.6	2.3	2.5	3.1	2.9	3.5	3.5
M-F	16.6	99.9	0.0	0.0	11.4	70.9	100.0	0.0	10.8	53.4	97.3	100.0
M-K	4.2	5.0	4.1	3.6	5.8	6.7	7.6	7.7	6.2	8.0	9.1	10.2
H-Z	3.8	3.0	2.6	2.2	4.6	3.6	3.5	3.0	4.3	4.2	3.7	3.8
Roy	4.9	5.2	4.7	4.8	5.2	4.9	5.1	5.1	4.8	5.2	5.3	4.7
MSL	4.9	8.2	12.7	17.0	5.8	10.7	16.3	21.2	6.4	14.0	20.0	24.6
R-O	4.9	7.0	9.3	10.6	4.8	7.0	9.9	11.5	4.5	7.2	9.4	11.7
Koz	6.2	7.6	7.4	8.1	5.2	5.6	6.5	6.4	5.3	4.9	5.7	5.6
Hawk	5.0	8.0	21.7	100.0	4.4	5.8	9.6	27.9	5.1	5.1	6.6	10.9
PRS	3.9	4.2	4.7	4.7	4.1	4.1	4.9	4.5	4.3	4.2	4.7	4.9
S	6.6	6.7	6.6	6.4	5.1	4.8	4.8	4.9	5.1	5.1	4.8	5.1
S_{ROB}	29.8	50.3	67.6	79.2	15.2	25.9	40.0	52.2	9.9	15.3	20.0	26.5

Note: Table entries are the percentages of the 10000 simulated data sets where the null hypothesis of multivariate normality was rejected at the $\alpha = 0.05$ level.
Skew = Mardia's test of multivariate skewness
Kurt = Mardia's test of multivariate kurtosis
M-F = Mardia-Foster omnibus test
M-K = Mardia-Kent omnibus test
H-Z = Henze-Zirkler test
Roy = Royston's test
MSL = Mudholkar-Srivastava-Lin test
R-O = Romeu-Ozturk test
Koz = Koziol's test
Hawk = Hawkins' test
PRS = Paulson-Roohan-Sullo test
S = Singh's test (classical)
S_{ROB} = Singh's test (robust)

Table 2

Empirical Power Against the Multivariate Normal Mixture Type 4 With Mixing Parameter = 0.9

MVN	n=25				n=50				n=100			
Test	p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5
Skew	7.2	7.9	8.4	8.0	8.7	10.2	10.6	10.2	9.5	12.7	14.2	14.0
Kurt	1.0	0.6	0.6	1.1	2.9	2.4	2.1	2.3	4.5	4.3	3.7	3.9
M-F	13.8	99.8	0.0	0.0	9.7	66.3	100.0	0.0	8.9	43.6	93.7	100.0
M-K	5.1	5.9	5.4	4.5	8.1	10.2	11.0	11.2	10.7	14.6	16.9	16.6
H-Z	4.2	3.1	2.7	2.7	4.0	4.0	3.7	3.6	4.8	4.6	4.6	4.4
Roy	6.5	6.1	12.8	14.4	5.2	6.4	6.2	8.1	5.7	5.8	6.2	6.5
MSL	7.1	12.9	24.8	37.5	5.6	11.5	18.9	27.7	6.7	12.5	20.3	28.6
R-O	4.7	6.9	9.6	11.3	4.8	7.4	9.3	12.1	5.1	7.4	10.0	11.2
Koz	6.1	6.1	6.4	7.5	5.3	5.3	5.6	5.9	5.9	5.4	4.9	5.0
Hawk	5.3	7.6	21.9	100.0	5.3	6.7	9.7	29.7	6.1	6.7	7.4	12.8
PRS	4.2	3.9	3.8	4.6	4.7	4.3	4.3	4.4	5.2	5.1	4.6	4.6
S	7.2	6.9	6.8	7.0	6.2	6.2	6.0	5.9	6.3	6.8	6.7	6.5
S_{ROB}	31.4	52.2	70.2	80.2	17.8	29.9	43.2	57.7	12.5	19.6	25.3	32.4

Note: Table entries are the percentages of the 10000 simulated data sets where the null hypothesis of multivariate normality was rejected at the $\alpha = 0.05$ level.

Skew = Mardia's test of multivariate skewness
Kurt = Mardia's test of multivariate kurtosis
M-F = Mardia-Foster omnibus test
M-K = Mardia-Kent omnibus test
H-Z = Henze-Zirkler test
Roy = Royston's test
MSL = Mudholkar-Srivastava-Lin test
R-O = Romeu-Ozturk test
Koz = Koziol's test
Hawk = Hawkins' test
PRS = Paulson-Roohan-Sullo test
S = Singh's test (classical)
S_{ROB} = Singh's test (robust)

Table 3

Empirical Power Against the Multivariate Normal Mixture Type 4 With Mixing Parameter = 0.788675

MVN	n=25				n=50				n=100			
Test	p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5
Skew	8.0	7.9	8.3	8.7	9.3	10.9	10.7	12.1	11.5	16.1	17.3	17.6
Kurt	1.2	0.5	0.7	1.0	2.6	2.5	1.8	2.4	3.3	3.9	3.3	3.5
M-F	14.4	99.8	0.0	0.0	9.1	61.7	99.9	0.0	7.7	35.2	90.8	100.0
M-K	5.1	5.5	5.3	4.6	7.9	10.2	10.2	11.4	10.1	15.0	16.6	17.3
H-Z	4.4	3.4	3.1	3.0	4.9	4.2	3.9	4.0	4.7	4.9	5.1	5.0
Roy	5.7	6.1	12.4	14.0	5.5	5.9	6.2	8.6	6.1	6.4	6.8	7.0
MSL	6.3	13.2	25.5	38.7	6.6	13.9	23.2	33.1	8.1	18.6	30.3	42.5
R-O	4.2	6.9	9.3	11.4	4.7	7.5	9.8	12.4	5.3	7.4	10.9	13.0
Koz	5.6	6.6	6.6	7.4	5.7	5.6	5.3	5.9	5.7	5.1	4.7	4.7
Hawk	5.0	7.9	21.6	100.0	5.5	6.4	9.9	29.6	5.5	6.4	7.5	12.3
PRS	4.2	3.8	3.9	4.6	4.9	4.5	4.4	4.4	4.7	4.9	4.6	4.7
S	7.5	6.8	6.7	6.5	5.4	5.9	5.2	5.5	5.4	6.2	5.8	5.7
S_{ROB}	31.3	53.3	69.3	80.5	16.9	29.9	41.5	56.5	11.5	18.0	23.7	30.5

Note: Table entries are the percentages of the 10000 simulated data sets where the null hypothesis of multivariate normality was rejected at the $\alpha = 0.05$ level.

Skew = Mardia's test of multivariate skewness
Kurt = Mardia's test of multivariate kurtosis
M-F = Mardia-Foster omnibus test
M-K = Mardia-Kent omnibus test
H-Z = Henze-Zirkler test
Roy = Royston's test
MSL = Mudholkar-Srivastava-Lin test
R-O = Romeu-Ozturk test
Koz = Koziol's test
Hawk = Hawkins' test
PRS = Paulson-Roohan-Sullo test
S = Singh's test (classical)
S_{ROB} = Singh's test (robust)

Table 4

Empirical Power Against the Multivariate Normal Mixture Type 4 With Mixing Parameter = 0.5

MVN	n=25				n=50				n=100			
Test	p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5
Skew	6.7	8.3	8.6	9.6	7.7	10.4	11.8	13.5	9.9	14.0	17.5	19.6
Kurt	0.8	0.6	0.5	0.8	2.0	2.3	2.3	2.8	3.0	4.0	4.2	4.2
M-F	13.9	99.8	0.0	0.0	9.2	61.4	99.9	0.0	8.2	37.3	92.3	100.0
M-K	4.3	5.5	5.8	5.5	6.8	9.8	11.8	13.4	8.5	14.1	17.5	21.3
H-Z	3.5	3.2	2.6	2.7	4.7	4.1	3.9	4.2	4.9	5.1	5.1	5.3
Roy	5.3	5.3	10.5	11.2	4.7	5.0	5.3	6.9	6.1	6.4	6.3	7.4
MSL	5.9	12.2	24.3	38.0	6.4	15.4	25.0	34.1	8.9	22.2	34.7	43.6
R-O	4.9	7.7	9.2	11.7	5.2	8.1	10.7	12.5	5.8	9.0	11.9	13.4
Koz	6.0	6.3	6.4	6.6	5.5	6.0	5.2	5.6	5.8	5.4	4.8	4.5
Hawk	4.8	7.6	22.4	100.0	4.7	6.5	10.2	31.7	5.2	6.2	7.5	14.5
PRS	4.0	3.7	3.7	4.0	4.5	4.7	4.2	4.7	4.5	5.0	4.6	5.1
S	6.9	6.6	7.3	7.4	5.3	6.3	5.7	6.2	5.2	6.4	6.8	7.5
S_{ROB}	30.9	51.5	70.8	82.2	15.8	30.1	44.1	59.4	11.1	18.9	27.0	36.6

Note: Table entries are the percentages of the 10000 simulated data sets where the null hypothesis of multivariate normality was rejected at the $\alpha = 0.05$ level.

Skew = Mardia's test of multivariate skewness
Kurt = Mardia's test of multivariate kurtosis
M-F = Mardia-Foster omnibus test
M-K = Mardia-Kent omnibus test
H-Z = Henze-Zirkler test
Roy = Royston's test
MSL = Mudholkar-Srivastava-Lin test
R-O = Romeu-Ozturk test
Koz = Koziol's test
Hawk = Hawkins' test
PRS = Paulson-Roohan-Sullo test
S = Singh's test (classical)
S_{ROB} = Singh's test (robust)

Table 5

Empirical Power Against the Multivariate Uniform Distribution

MVN	n=25				n=50				n=100			
Test	p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5
Skew	0.3	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Kurt	89.0	95.8	97.2	96.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
M-F	92.1	99.3	67.6	37.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
M-K	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.1	0.0	0.0
H-Z	100.0	99.8	72.3	39.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Roy	100.0	66.5	50.7	40.3	100.0	100.0	97.7	82.1	100.0	100.0	100.0	100.0
MSL	97.8	38.5	29.6	37.7	100.0	96.1	61.0	42.3	100.0	100.0	98.3	89.9
R-O	85.3	35.8	20.9	18.5	99.9	78.3	48.6	35.8	100.0	99.5	88.4	76.4
Koz	100.0	100.0	99.9	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Hawk	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
PRS	99.9	99.9	99.9	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
S	39.4	17.0	13.5	13.5	76.5	38.3	21.5	14.5	99.6	74.1	56.2	37.9
S_{ROB}	39.8	17.9	15.1	16.3	76.5	38.3	21.5	14.6	99.6	74.1	56.2	37.9

Note: Table entries are the percentages of the 10000 simulated data sets where the null hypothesis of multivariate normality was rejected at the $\alpha = 0.05$ level.

Skew = Mardia's test of multivariate skewness
Kurt = Mardia's test of multivariate kurtosis
M-F = Mardia-Foster omnibus test
M-K = Mardia-Kent omnibus test
H-Z = Henze-Zirkler test
Roy = Royston's test
MSL = Mudholkar-Srivastava-Lin test
R-O = Romeu-Ozturk test
Koz = Koziol's test
Hawk = Hawkins' test
PRS = Paulson-Roohan-Sullo test
S = Singh's test (classical)
S_{ROB} = Singh's test (robust)

Table 6

Empirical Power Against the Multivariate t Distribution

MVN Test	n=25 p=2	n=25 p=3	n=25 p=4	n=25 p=5	n=50 p=2	n=50 p=3	n=50 p=4	n=50 p=5	n=100 p=2	n=100 p=3	n=100 p=4	n=100 p=5
Skew	5.9	6.4	6.5	5.8	6.1	5.8	6.3	6.9	5.8	6.5	6.5	6.0
Kurt	0.9	0.6	0.7	1.2	2.0	1.7	1.8	2.7	3.2	3.2	3.1	3.6
M-F	15.7	89.9	0.0	0.0	10.4	70.6	100.0	0.0	10.9	53.1	97.1	100.0
M-K	4.1	4.6	4.4	3.1	6.0	6.7	7.6	7.8	6.7	8.5	9.2	9.6
H-Z	3.9	3.0	2.6	2.3	4.5	3.5	3.5	3.1	5.2	4.3	4.2	3.6
Roy	5.4	5.0	5.0	4.6	5.1	5.1	5.1	7.1	5.2	5.5	5.5	5.5
MSL	4.7	8.3	12.7	17.3	5.5	10.9	16.6	26.7	6.5	13.3	20.0	27.0
R-O	5.0	6.8	8.9	11.4	5.0	6.8	9.3	11.7	5.1	7.7	9.6	11.5
Koz	6.0	7.1	7.6	8.0	5.0	6.0	5.7	6.8	5.4	5.4	5.1	5.4
Hawk	4.6	7.7	21.6	100.0	4.3	5.7	8.6	28.6	5.0	5.6	6.4	10.9
PRS	3.8	4.1	4.6	4.8	4.0	4.3	4.0	4.7	4.2	4.7	4.5	4.6
S	6.6	6.8	6.3	6.6	5.3	5.0	5.1	5.1	4.9	5.4	4.8	5.6
S_{ROB}	29.5	50.9	68.3	78.9	15.0	26.1	39.4	53.1	9.9	14.9	20.2	27.4

Note: Table entries are the percentages of the 10000 simulated data sets where the null hypothesis of multivariate normality was rejected at the $\alpha = 0.05$ level.

Skew = Mardia's test of multivariate skewness
Kurt = Mardia's test of multivariate kurtosis
M-F = Mardia-Foster omnibus test
M-K = Mardia-Kent omnibus test
H-Z = Henze-Zirkler test
Roy = Royston's test
MSL = Mudholkar-Srivastava-Lin test
R-O = Romeu-Ozturk test
Koz = Koziol's test
Hawk = Hawkins' test
PRS = Paulson-Roohan-Sullo test
S = Singh's test (classical)
S_{ROB} = Singh's test (robust)

Table 7

Empirical Power Against the Multivariate Cauchy Distribution

MVN	n=25				n=50				n=100			
Test	p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5
Skew	6.3	6.4	6.7	6.4	5.8	6.3	6.6	6.6	4.9	5.8	6.1	6.1
Kurt	0.8	0.6	0.6	0.9	1.9	1.6	2.0	2.3	3.0	2.9	3.0	3.9
M-F	15.1	99.9	0.0	0.0	10.9	69.7	100.0	0.0	10.6	53.3	97.3	100.0
M-K	4.3	4.6	4.5	3.6	5.9	6.8	7.9	8.0	6.5	8.2	8.9	10.1
H-Z	3.7	3.0	2.7	2.6	4.2	3.3	3.8	3.3	4.9	4.4	4.0	3.5
Roy	4.9	5.1	5.0	14.1	5.5	5.0	5.2	7.0	5.0	5.1	5.0	6.0
MSL	5.0	8.6	12.7	34.8	5.9	10.8	19.6	26.6	6.6	12.7	19.5	26.3
R-O	4.7	6.9	8.8	11.0	5.1	6.9	8.9	11.6	4.9	6.6	10.4	11.5
Koz	6.0	7.3	6.9	8.1	5.1	5.8	6.1	6.8	5.7	5.0	5.1	5.8
Hawk	4.7	7.4	21.7	100.0	4.5	5.6	9.3	28.1	5.4	5.3	6.2	10.0
PRS	3.9	4.0	3.9	5.1	4.2	4.1	4.5	4.6	4.7	4.3	4.2	4.8
S	6.7	6.8	6.7	6.8	4.8	4.7	5.1	5.2	5.0	5.2	4.5	4.8
S_{ROB}	29.3	50.8	67.6	79.0	15.3	26.4	39.2	52.8	9.7	15.5	19.6	26.9

Note: Table entries are the percentages of the 10000 simulated data sets where the null hypothesis of multivariate normality was rejected at the $\alpha = 0.05$ level.

Skew = Mardia's test of multivariate skewness
Kurt = Mardia's test of multivariate kurtosis
M-F = Mardia-Foster omnibus test
M-K = Mardia-Kent omnibus test
H-Z = Henze-Zirkler test
Roy = Royston's test
MSL = Mudholkar-Srivastava-Lin test
R-O = Romeu-Ozturk test
Koz = Koziol's test
Hawk = Hawkins' test
PRS = Paulson-Roohan-Sullo test
S = Singh's test (classical)
S_{ROB} = Singh's test (robust)

Table 8

Empirical Power Against the Multivariate Chi-Square Distribution

MVN Test	n=25 p=2	n=25 p=3	n=25 p=4	n=25 p=5	n=50 p=2	n=50 p=3	n=50 p=4	n=50 p=5	n=100 p=2	n=100 p=3	n=100 p=4	n=100 p=5
Skew	99.3	99.7	99.8	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Kurt	70.0	74.7	73.7	71.9	96.5	98.6	99.3	99.5	100.0	100.0	100.0	100.0
M-F	51.7	94.4	3.9	0.3	96.4	97.3	98.8	62.0	100.0	100.0	100.0	100.0
M-K	93.7	96.2	96.8	96.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
H-Z	99.6	99.8	99.8	99.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Roy	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
MSL	99.8	99.2	98.8	97.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
R-O	99.5	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Koz	78.4	82.6	80.6	78.1	98.8	99.4	99.7	99.8	100.0	100.0	100.0	100.0
Hawk	81.0	91.3	96.8	100.0	98.9	99.6	99.9	100.0	100.0	100.0	100.0	100.0
PRS	75.6	81.4	82.3	82.1	98.6	99.3	99.7	99.8	100.0	100.0	100.0	100.0
S	82.8	83.5	80.5	77.0	97.0	97.8	97.9	97.7	99.9	100.0	100.0	100.0
S_{ROB}	96.0	99.3	99.7	99.9	98.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0

Note: Table entries are the percentages of the 10000 simulated data sets where the null hypothesis of multivariate normality was rejected at the $\alpha = 0.05$ level.

Skew = Mardia's test of multivariate skewness
Kurt = Mardia's test of multivariate kurtosis
M-F = Mardia-Foster omnibus test
M-K = Mardia-Kent omnibus test
H-Z = Henze-Zirkler test
Roy = Royston's test
MSL = Mudholkar-Srivastava-Lin test
R-O = Romeu-Ozturk test
Koz = Koziol's test
Hawk = Hawkins' test
PRS = Paulson-Roohan-Sullo test
S = Singh's test (classical)
S_{ROB} = Singh's test (robust)

Table 9

Empirical Power Against the Multivariate Lognormal Distribution

MVN	n=25				n=50				n=100			
Test	p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5
Skew	98.8	99.7	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Kurt	82.5	90.3	93.7	95.8	99.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
M-F	67.1	100.0	20.7	6.3	98.5	99.6	100.0	95.3	100.0	100.0	100.0	100.0
M-K	95.2	98.0	99.3	99.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
H-Z	99.1	99.7	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Roy	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
MSL	99.1	96.3	95.6	80.8	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0
R-O	97.1	98.9	99.4	99.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Koz	82.6	92.3	94.8	96.6	99.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Hawk	87.2	97.1	99.6	100.0	99.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0
PRS	83.0	92.9	95.8	97.7	99.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0
S	88.8	92.1	93.4	93.1	98.8	99.7	100.0	100.0	100.0	100.0	100.0	100.0
S_{ROB}	94.7	99.6	100.0	100.0	99.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: Table entries are the percentages of the 10000 simulated data sets where the null hypothesis of multivariate normality was rejected at the $\alpha = 0.05$ level.

Skew = Mardia's test of multivariate skewness
Kurt = Mardia's test of multivariate kurtosis
M-F = Mardia-Foster omnibus test
M-K = Mardia-Kent omnibus test
H-Z = Henze-Zirkler test
Roy = Royston's test
MSL = Mudholkar-Srivastava-Lin test
R-O = Romeu-Ozturk test
Koz = Koziol's test
Hawk = Hawkins' test
PRS = Paulson-Roohan-Sullo test
S = Singh's test (classical)
S_{ROB} = Singh's test (robust)

Table 10

Empirical Power Against the Knintchine Distribution

MVN Test	n=25 p=2	n=25 p=3	n=25 p=4	n=25 p=5	n=50 p=2	n=50 p=3	n=50 p=4	n=50 p=5	n=100 p=2	n=100 p=3	n=100 p=4	n=100 p=5
Skew	0.3	0.2	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Kurt	1.8	9.5	18.1	24.8	60.5	78.3	83.7	87.6	99.7	99.9	100.0	100.0
M-F	30.3	98.6	0.2	0.1	82.2	96.0	100.0	6.7	99.8	100.0	100.0	100.0
M-K	0.1	0.1	0.2	0.1	0.0	0.0	0.0	0.0	3.1	0.1	0.0	0.0
H-Z	12.7	10.5	9.2	7.3	53.2	49.9	44.0	37.2	94.8	94.8	93.8	90.5
Roy	20.0	15.6	12.5	33.4	84.7	80.8	78.8	79.8	100.0	100.0	100.0	100.0
MSL	19.8	22.0	24.4	49.1	80.5	71.0	69.6	70.0	100.0	99.7	99.3	98.9
R-O	28.1	36.1	41.0	45.7	64.9	77.1	83.4	88.0	96.2	99.0	99.7	100.0
Koz	54.5	59.6	58.1	57.9	89.0	92.4	92.9	93.0	99.8	100.0	100.0	100.0
Hawk	33.2	45.1	69.2	100.0	77.8	85.1	89.2	96.9	99.5	99.8	99.9	100.0
PRS	36.4	41.9	44.5	47.2	81.1	86.9	88.3	89.6	99.5	99.9	100.0	100.0
S	21.7	16.0	13.3	11.7	45.1	30.7	22.5	17.8	90.2	76.0	61.4	51.2
S_{ROB}	26.6	28.8	36.2	43.2	45.0	32.6	27.7	28.7	88.9	74.8	60.7	51.6

Note: Table entries are the percentages of the 10000 simulated data sets where the null hypothesis of multivariate normality was rejected at the $\alpha = 0.05$ level.

Skew = Mardia's test of multivariate skewness
Kurt = Mardia's test of multivariate kurtosis
M-F = Mardia-Foster omnibus test
M-K = Mardia-Kent omnibus test
H-Z = Henze-Zirkler test
Roy = Royston's test
MSL = Mudholkar-Srivastava-Lin test
R-O = Romeu-Ozturk test
Koz = Koziol's test
Hawk = Hawkins' test
PRS = Paulson-Roohan-Sullo test
S = Singh's test (classical)
S_{ROB} = Singh's test (robust)

Table 11

Empirical Power Against the Generalized Exponential Power Distribution

MVN	n=25				n=50				n=100			
Test	p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5	p=2	p=3	p=4	p=5
Skew	5.6	6.2	7.0	6.7	4.0	4.9	5.5	6.2	2.9	3.9	4.5	4.8
Kurt	0.9	0.5	0.5	0.6	1.2	1.0	1.1	1.7	1.3	1.6	1.8	2.0
M-F	18.3	99.9	0.0	0.0	12.2	73.6	100.0	0.0	10.7	57.5	97.9	100.0
M-K	3.5	4.0	4.2	3.6	3.5	4.1	4.8	5.8	2.3	3.6	4.7	5.8
H-Z	4.7	3.5	2.8	2.8	5.4	5.0	4.0	3.9	7.8	6.9	6.3	5.9
Roy	5.5	5.6	9.7	10.8	4.6	4.6	4.7	6.1	5.6	6.0	5.1	5.5
MSL	5.8	11.8	21.9	35.5	5.7	11.8	17.6	26.3	7.3	13.9	20.1	27.4
R-O	5.6	7.4	10.1	12.2	8.1	11.1	13.2	15.9	13.7	20.6	25.0	29.0
Koz	6.5	6.3	6.9	6.9	8.8	7.0	6.3	6.8	16.4	11.0	8.4	7.1
Hawk	9.2	11.0	23.7	100.0	14.1	12.5	12.9	28.1	24.7	18.6	14.7	16.2
PRS	6.5	5.9	5.4	4.9	12.4	9.5	7.4	6.5	22.3	16.4	12.1	9.9
S	4.5	5.5	6.2	6.3	2.7	3.4	3.8	4.2	3.1	3.5	4.0	4.3
S_{ROB}	27.0	47.8	67.5	78.1	9.0	18.1	30.7	45.2	3.6	6.5	9.6	15.1

Note: Table entries are the percentages of the 10000 simulated data sets where the null hypothesis of multivariate normality was rejected at the $\alpha = 0.05$ level.

Skew = Mardia's test of multivariate skewness
Kurt = Mardia's test of multivariate kurtosis
M-F = Mardia-Foster omnibus test
M-K = Mardia-Kent omnibus test
H-Z = Henze-Zirkler test
Roy = Royston's test
MSL = Mudholkar-Srivastava-Lin test
R-O = Romeu-Ozturk test
Koz = Koziol's test
Hawk = Hawkins' test
PRS = Paulson-Roohan-Sullo test
S = Singh's test (classical)
S_{ROB} = Singh's test (robust)

AERA

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM030864

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: A COMPARISON OF THE POWER OF CLASSICAL AND NEWER TESTS OF MULTIVARIATE NORMALITY

Author(s): CHRISTOPHER J. MECKLIN AND DANIEL J. MUNDFROM

Corporate Source:

Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 [X]	Level 2A [ ]	Level 2B [ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature: Christopher J. Mecklin

Printed Name/Position/Title: Christopher J. Mecklin

Organization/Address: Department of Applied Statistics and Research Methods
University of Northern Colorado
Greeley CO 80639

Telephone: (970) 304-1352

FAX:

E-Mail Address: meck1206@blue.unco.edu

Date: 4/27/00

(over)

ERIC®

Clearinghouse on Assessment and Evaluation

University of Maryland
1129 Shriver Laboratory
College Park, MD 20742-5701

Tel: (800) 464-3742
(301) 405-7449
FAX: (301) 405-8134
ericae@ericae.net
http://ericae.net

March 2000

Dear AERA Presenter,

Congratulations on being a presenter at AERA. The ERIC Clearinghouse on Assessment and Evaluation would like you to contribute to ERIC by providing us with a written copy of your presentation. Submitting your paper to ERIC ensures a wider audience by making it available to members of the education community who could not attend your session or this year's conference.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed, electronic, and internet versions of *RIE*. The paper will be available **full-text, on demand through the ERIC Document Reproduction Service** and through the microfiche collections housed at libraries around the world.

We are gathering all the papers from the AERA Conference. We will route your paper to the appropriate clearinghouse and you will be notified if your paper meets ERIC's criteria. Documents are reviewed for contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality. You can track our processing of your paper at **http://ericae.net.**

To disseminate your work through ERIC, you need to sign the reproduction release form on the back of this letter and include it with **two** copies of your paper. You can drop of the copies of your paper and reproduction release form at the ERIC booth (223) or mail to our attention at the address below. **If you have not submitted your 1999 Conference paper please send today or drop it off at the booth with a Reproduction Release Form.** Please feel free to copy the form for future or additional submissions.

Mail to: AERA 2000/ERIC Acquisitions
The University of Maryland
1129 Shriver Lab
College Park, MD 20742

Sincerely,

Lawrence M. Rudner

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

ERIC/AE is a project of the Department of Measurement, Statistics and Evaluation at the College of Education, University of Maryland.